Acoustic Metadata Management and Transparent Access to Networked Oceanographic Data Sets

Marie A. Roch
Dept. of Computer Science
San Diego State University
5500 Campanile Drive
San Diego, CA 92182-7720
phone: (619) 594-5830 fax: (619) 594-6746 e-mail: marie.roch@sdsu.edu

Simone Baumann-Pickering & John A. Hildebrand
Scripps Institution of Oceanography
University of California at San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0205
phone: (858) 534-7280 fax: (858) 534-6849 e-mail: sbaumann@ucsd.edu
phone: (858) 534-4069 fax: (858) 534-6849 e-mail: jhildebrand@ucsd.edu

Catherine L. Berchok
Alaska Fisheries Science Center
National Marine Fisheries Service
7600 Sand Point Way N.E., Building 4
Seattle, Washington 98115-6349
phone: (206) 526-6331 fax: (206) 526-6615 e-mail: Catherine.Berchok@noaa.gov

Erin M. Oleson
Pacific Islands Fisheries Science Center
National Marine Fisheries Service
1601 Kapiolani Blvd, Ste. 1110
Honolulu, Hawaii 996814
phone: (808) 944-2172 fax: (808) 941-0307 e-mail: Erin.Oleson@noaa.gov

Sofie M. Van Parijs
Northeast Fisheries Science Center
National Marine Fisheries Service
166 Water Street, Woods Hole, MA 02543
phone: (508) 495-2119 fax: (508) 495-2258 e-mail: sofie.vanparijs@noaa.gov

Award Numbers: N00014-11-1-0697, N0001414IP20032,
N0001413IP20029, & N0001412IP20059
http://tethys.sdsu.edu

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
30 SEP 2014		**00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Acoustic Metadata Management and Transparent Access to Networked Oceanographic Data Sets	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **San Diego State University,Dept. of Computer Science,5500 Campanile Drive,San Diego,CA,92182-7720**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

LONG-TERM GOALS

Passive acoustic monitoring (PAM) of sound-producing animals (e.g. marine mammals) is producing multi-year data sets in many locations. Such long-term data sets allow the investigation of questions related to the effects of environmental processes and anthropogenic activities on these animals at varying temporal-spatial scales. Preservation of acoustic metadata in a standardized format can prevent the loss of detailed information that is not captured in scientific publications and permits the analysis of decadal-scale data sets and beyond.

The need for standardization is juxtaposed with the need to include new information for specific research questions that evolve over time. This work proposes a schema for organizing acoustic metadata that addresses many of the problems associated with the retention of metadata from long-term passive acoustic data sets and enables the incorporation of research-question specific details without the need for changes to the schema or its implementation.

The long-term goals of this effort are to produce community data standards and software capable of organizing and archiving acoustic metadata associated with passive acoustic monitoring of marine mammals. The software provides a reference implementation of the community standards in a scientific workbench setting. The workbench enables researchers to integrate queries for both biological and physical data by providing mediation services to support the retrieval of Internet-available data sets in a manner similar to those used for querying the acoustic data.

The data are accessible from a variety of languages used by the scientific community for analysis and modeling. The Tethys server at Scripps now has data sets that cover over nine years, and published and submitted articles are demonstrating the benefits to the preservation of consistent metadata from long time-series data.

OBJECTIVES

The objectives of this effort are to produce:

1. A database which can flexibly store multiple types of acoustic metadata derived from a variety of acoustic platforms, both stationary and mobile.
2. Standardization of methods to make the data repositories useful to the passive acoustic monitoring community.
3. Access to network available data products in a standard manner (e.g. ephemeris).
4. Secure access on network platforms using industry standard security protocols.
5. Query and visualization primitives in selected analysis and modeling languages (e.g. Matlab, R) for efficient manipulation of spatial-temporal data.
6. Demonstration projects to show the value of the database as a scientific workbench component.

APPROACH

1) Technical approach

The Tethys Metadata Workbench enables researchers to organize, store and most importantly query information derived from passive acoustic monitoring (PAM). Due to the large number of acquisition platforms, types of detection effort, etc., structuring these data is a complicated semi-structured task and traditional databases do not meet the needs of PAM users. By compiling a large team of PAM users who work on a global scale, we are defining data standards that are likely to meet the needs of the PAM community in general. Networking capabilities provide the ability to share data and export summary data to OBIS-SEAMAP. In addition, this effort provides users with access to online physical oceanography databases using a single interface. The server can be accessed directly through a web services framework or through libraries that provide access tools in a variety of computer languages used by the scientific community.

2) Key Personnel

Dr. Marie A. Roch (San Diego State University) is the project manager and administrator for this project. She also takes the lead for software development.

Dr. John A Hildebrand (Scripps Institution of Oceanography (SIO)) is the project manager for the subaward to SIO.

Drs. Simone Baumann-Pickering (Scripps Institution of Oceanography), Catherine L. Berchok (NOAA Alaska Fisheries Science Center (AFSC)), Erin M. Oleson (NOAA Pacific Island Fisheries Science Center (PIFSC)), Melissa Soldevilla (NOAA Southeast Fisheries Science Center (SEFSC)), and Sofie Van Parijs (NOAA Northeast Fisheries Science Center (NEFSC)), all represent data providers who will be using the database and are integrally involved in the operational specification, requirements, and testing.

Dr. Simone Baumann-Pickering is providing the lead on habitat modeling, and Dr. Sofie Van Parijs is the project manager for NOAA as well as the lead on data standardization.

3) Work plans for the upcoming year

Work plans for the final months of this project include completion of the R language library and the security model. Several papers have been submitted, and any needed revisions will be made. We will have our final team workshop in December which will focus on effective use of Tethys. We will continue to publicize Tethys, such as a presentation at the upcoming Indianapolis Acoustical Society meeting.

WORK COMPLETED

Development this year focused on six areas:

1. Refinement of schema (data model) to represent metadata.
2. Development of a client library for the R programming language.
3. Server improvements to performance, security, and standardization.

4. A push amongst NOAA participants to move forward with making Tethys part of their standard practices.
5. We worked with Ei Fujioka of Duke University to develop methods to publish summary results from Tethys to the Ocean Biogeographic Observing System – Spatial Ecological Analysis of Megavertebrate Populations (OBIS-SEAMAP).
6. Two manuscripts were submitted, and a third was published.

The schema were refined to permit the incorporation of tracklines along with deployments. Previously, it had been assumed that these would be stored elsewhere, but the ability to reason spatially about moving platforms within the context of the Tethys Metadata Workbench was compelling and we thus modified the schema. A number of other minor improvements to schema were made.

Two of the most popular languages for writing bioacoustics analysis programs are Matlab and R. We continue to enhance our Matlab library and have developed a rudimentary R library that is not yet ready for use by non-experts.

Several major improvements were made to the server this year. We transitioned to using a 64 bit architecture which greatly improves our ability to handle large datasets. We also added indices that reduce the amount of time required for many classes of queries. The communication model between the server and clients was changed to use a so-called RESTful architecture (Fielding, 2000) which is a standard method to provide web services and will allow greater interoperability between Tethys and other systems that may wish to query Tethys servers. It also provides compression of data across a network, resulting in faster transmission of data between the server and clients. A cache was added for queries to remote data products (e.g. NASA Ocean Color). When scientists are analyzing data, the same query is often repeated multiple times as they develop their code base. The cache reduces the load on the remote repository servers and equally importantly dramatically speeds subsequent queries. We also refined encrypted data transmission and wrote an encrypted data user guide. This is a necessary step for user account control that is currently being developed.

A dedicated person (Dana Gerlach) has been assigned at NOAA Northeast Fisheries Science Center to develop Tethys expertise and import acoustic metadata. Ms. Gerlach has been working with members of the development team and the expertise is being transferred to other NOAA Fisheries Science Centers.

An article discussing echolocation signals of beaked whales (work expanded from our featured project in our 2011 report) was published in PLOS-ONE. Three manuscripts were submitted, one examining spatial-temporal distributions of blue and fin whales in the Southern California Bight, another examining the impact of equipment and site variability on the acoustic classification of odontocetes, and a final one focusing on the Tethys schema.

RESULTS

Advances made this year have standardized data exchange, permitting the Tethys Workbench to interoperate with other programs across the Internet. Improvements have been made both in speed and the ability to represent large documents such as those produced by automated detectors. The development of an extensible markup language (XML) generation interface for detections permitted

the coupling of a blue and fin whale automated detectors, resulting in the addition of millions of call-level detections to a Tethys Workbench server. The ability to export to OBIS-SEAMAP means that summaries of work undertaken by individual laboratories can now be shared with the broader community. We demonstrated this with the export of several million detection records, a subset of which are being released publicly (many of the records are related to unpublished automated detector work and will be made public once the work is published).

One manuscript on world-wide beaked whale descriptions was published (Baumann-Pickering *et al.*, 2014), and three more have been submitted (Roch *et al.*, submitted-a; Roch *et al.*, submitted-b; Širović *et al.*, submitted). Roch et al. (submitted-a) is a description of the Tethys schema. Širović et al. (submitted) is a spatiotemporal analysis of blue and fin whale detections that analyzes sixteen different sites across the Southern California Bight between the years 2006 and 2012, and takes into account propagation modeling at each site. Finally, Roch et al. (submitted-b) is an echolocation click classification study that examined the effects of equipment and site variability across a range of instruments and years in the Southern California Bight. Analyses conducted in the three studies were greatly simplified by the use of the Tethys Metadata Workbench.

As in previous years, we will focus on some of the advances that have been enabled using Tethys rather than the Tethys system itself by examining the research of Širović et al. (submitted) who expanded their work this year to account for site-specific distances in detectability. These differences can vary dramatically for different locations as illustrated by detection-range shading for three of the study sites (Figure 1). Detection distance was estimated using range-dependent acoustic transmission loss models (RAM) in the ESME 2012 Workbench (Mountain *et al.*, 2012), demonstrating the synergy between these two ONR projects. The detection area was used to normalize the acoustic counts based on the area over which the counts were detected. The study integrates a wide variety of technologies to perform analysis of automated detections over a seven year period across a local region (The Southern California Bight) while recognizing that different locations can have dramatically different detection ranges, biasing the simple counts obtained at each site.

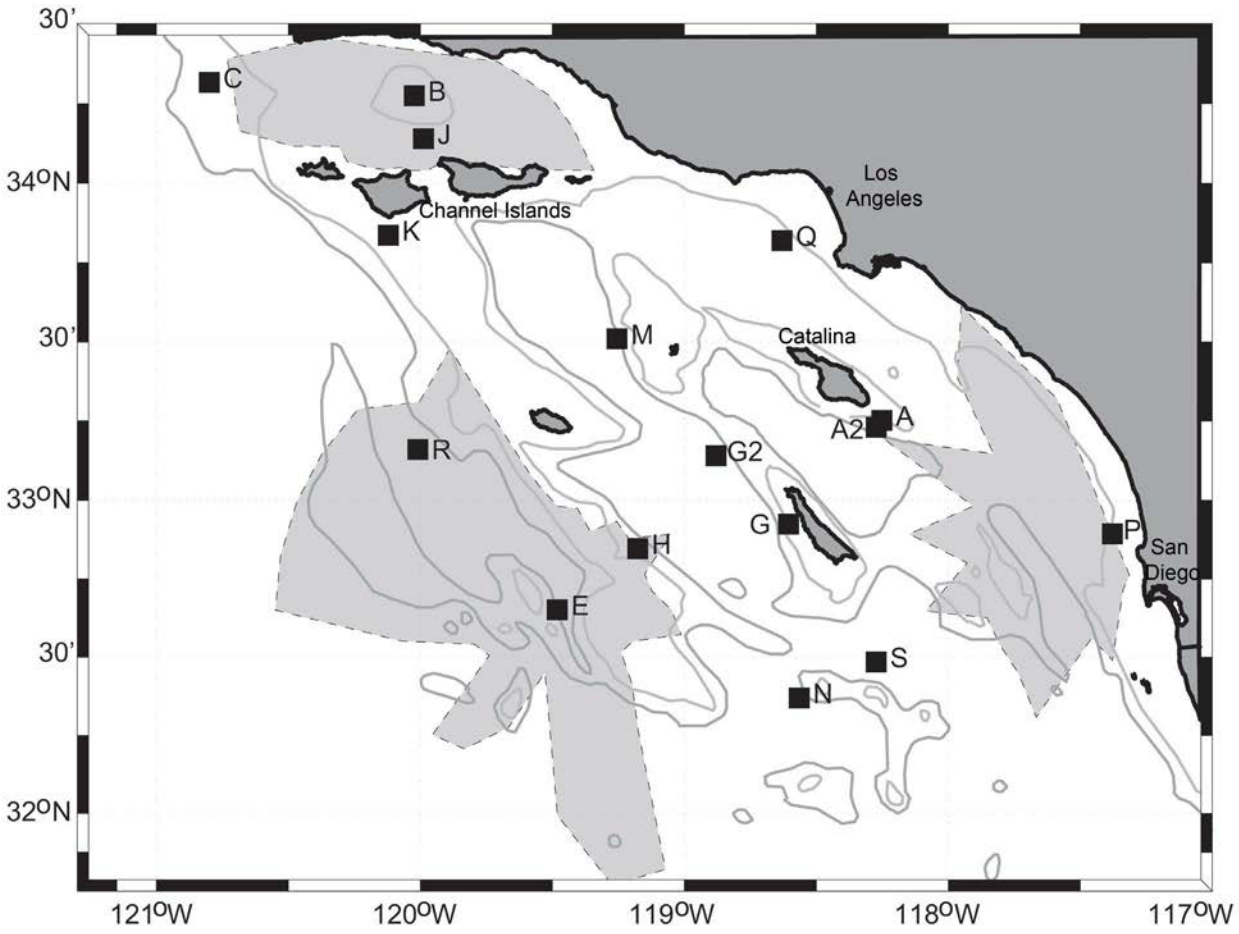

Figure 1 (Širović et al., submitted) – Locations of HARP deployment sites in the Southern California Bight. Examples of detection areas for blue whale calls obtained by propagation modeling along radials to a common detection threshold are shown for sites B, E, and P as gray shaded areas within a broken line. Lighter grey line marks 500 m and darker line is 1,000 m bathymetry contour.

Seasonal patterns can be seen in the normalized detections (Figure 2), and the normalization process permits call detections to be used as a proxy for calling animal abundance at specific sites (as always, silent animals will not be detected by passive acoustic monitoring). While complete details of the paper are beyond the scope of this report, Širović et al. (submitted) examine seasonal and spatial use of the Southern California Bight. The Tethys Workbench tracked instrument deployment locations, call detection effort and associated detections, and its query facility let the data be easily manipulated into forms required to answer specific questions. Integration of the new XML generation library enabled the automated detectors to generate Tethys ready data that could be easily added to the database without any conversion.

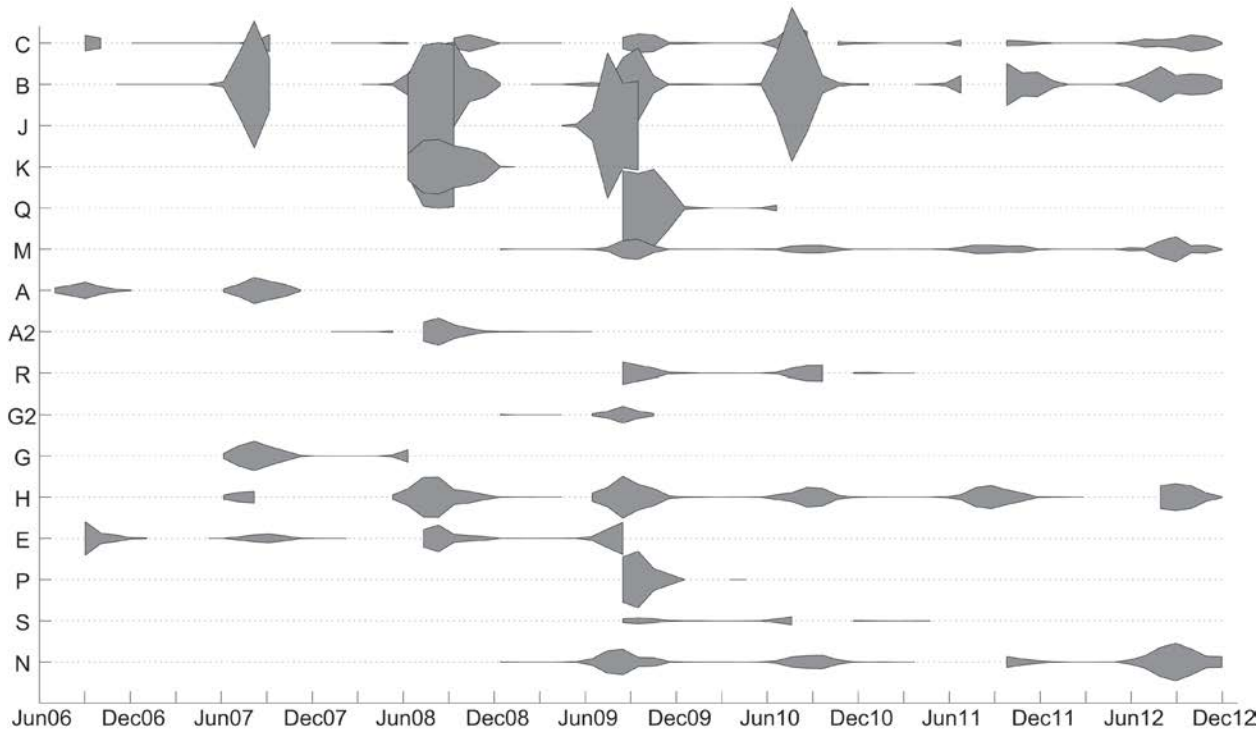

Figure 2 (Širović et al., submitted) – Detection of normalized blue whale B call counts (averaged monthly) at each site in the Southern California Bight. Sites are arranged, to the maximum extent possible, from the northernmost sites at the top towards the southernmost sites at the bottom. Size of the patch represents the call detection rate normalized by the area in which a call is likely to be detected. Dotted lines are periods with no data at that site and straight lines denote periods with recording but no detected calls.

IMPACT/APPLICATIONS

The most mature Tethys server node currently has over 4.5 million detections with an acoustic effort that spans nine years at sites in the Pacific and Atlantic Oceans, The Bering Sea, and The Gulf of Mexico. This includes data for many US Navy sponsored deployments.

TRANSITIONS

Metadata for all US Navy sponsored high-frequency acoustic recording package (HARP) analysis is currently being stored using the Tethys Metadata Workbench. In addition, NOAA Northeast Fisheries Science Center has begun importing more substantial parts of their passive acoustic metadata and is providing assistance to other Fisheries Science Centers to follow suit. The Tethys metatdata used to describe deployments are being considered for a trial of archiving NOAA acoustic data with the National Geographic Data Center.

RELATED PROJECTS

ONR N00014-13-IP20051– Advanced Methods for Passive Acoustic Detection, Classification, and Localization of Marine Mammals. PI Jonathan Klay, Dave Mellinger, Dave Moretti, Steve Martin

and Marie A. Roch. Some of the work in this grant makes use of Tethys and has overlapping key personnel.

N00014-12-1-0273 – Modeling of Habitat and Foraging Behavior of Beaked Whales in the Southern California Bight, PI John Hildebrand, Simone Baumann-Pickering – The work performed in this grant makes use of Tethys and has overlapping key personnel.

N000141210904 – Blue and fin whale habitat modeling from long-term year-round passive acoustic data from the Southern California Bight, PI John Hildebrand, Ana Širović. – The work performed in this grant makes use of Tethys and has overlapping key personnel.

N000141310641 – ESME workbench enhancements – PI David Mountain – ESME provides acoustic modeling, simulated animal movements, and environmental data visualization.

NSF-OCE-11-38046 – OBIS-SEAMAP, PI Patrick N. Halpin – OBIS-SEAMAP collects visual and acoustic detection information for marine mammals, sea birds, and sea turtles. We have worked with Ei Fujioka to integrate acoustic detections into their platform to permit transfer of data summaries from Tethys to OBIS-SEAMAP.

REFERENCES

Baumann-Pickering, S., Roch, M. A., Brownell Jr, R. L., Simonis, A. E., McDonald, M. A., Solsona-Berga, A., Oleson, E. M., Wiggins, S. M., and Hildebrand, J. A. (2014). "Spatio-Temporal Patterns of Beaked Whale Echolocation Signals in the North Pacific," Plos One 9(1), pp. 17, doi:10.1371/journal.pone.0086072.g001.

Fielding, R. T. (2000). "Architectural Styles and the Design of Network-based Software Architectures," *Computer Science*, The University of CA, Irvine, Irvine, CA, pp. 162.

Mountain, D., Anderson, D., Brughera, A., Cross, M., Houser, D., Musleh, N., Porter, M., and Siderius, M. (2012). "The ESME Workbench: Simulating the Impact of Anthropogenic Sound on Marine Mammals," in *The Effects of Noise on Aquatic Life*, edited by A. Popper, and A. Hawkins (Springer New York), pp. 217-219.

Roch, M. A., Batchelor, H., Baumann-Pickering, S., Berchock, C. L., Cholewiak, D., Fujioka, E., Garland , E. C., Herbert, S., Hildebrand, J. A., Munger, L. M., Oleson, E. M., Van Parijs, S. M., Risch, D., and Širović , A. (submitted-a). "Tethys: Metadata Management for Bioacoustics."

Roch, M. A., Stinner-Sloan, J., and Baumann-Pickering, S. (submitted-b). "The Effects of Site and Equipment Variation on Species Identification of Delphinids from Their Echolocation Clicks."

Širović , A., Rice, A., Chou, E., Hildebrand, J. A., and Roch, M. A. (submitted). "Seven years of blue and fin whale call abundance in Southern California."